Mining MOOC Lecture Transcripts to Construct Concept Dependency Graphs

Fareedah ALSaad*
University of Illinois at Urbana-Champaign
Urbana-Champaign, USA
alsaad2@illinois.edu

Assma Boughoula
University of Illinois at Urbana-Champaign
Urbana-Champaign, USA
boughou1@illinois.edu

Chase Geigle
University of Illinois at Urbana-Champaign
Urbana-Champaign, USA
geigle1@illinois.edu

Hari Sundaram
University of Illinois at Urbana-Champaign
Urbana-Champaign, USA
hs1@illinois.edu

ChengXiang Zhai
University of Illinois at Urbana-Champaign
Urbana-Champaign, USA
czhai@illinois.edu

ABSTRACT

This paper addresses the question of identifying a concept dependency graph for a MOOC through unsupervised analysis of lecture transcripts. The problem is important: extracting a concept graph is the first step in helping students with varying preparation to understand course material. The problem is challenging: instructors are unaware of the student preparation diversity and may be unable to identify the right resolution of the concepts, necessitating costly updates; inferring concepts from groups suffers from polysemy; the temporal order of concepts depends on the concepts in question. We propose innovative unsupervised methods to discover a directed concept dependency within and between lectures. Our main technical innovation lies in exploiting the temporal ordering amongst concepts to discover the graph. We propose two measures—the Bridge Ensemble Measure and the Global Direction Measure—to infer the existence and the direction of the dependency relations between concepts. The bridge ensemble measure identifies concept overlap between lectures, determines concept co-occurrence within short windows, and the lecture where concepts occur first. The global direction measure incorporates time directly by analyzing the concept time ordering both globally and within lectures. Experiments over real-world MOOC data show that our method outperforms the baseline in both AUC and precision/recall curves.

Keywords

Concept Dependency Graph, Temporal Order, Bridge Ensemble Measure, Global Direction Measure, Edge Direction, Edge Existence.

*King AbdulAziz University, Jeddah, Saudi Arabia.

1. INTRODUCTION

This paper presents two methods to identify extant concept relationships in lectures from a Massive Open Online Course (MOOC).

The problem of concept relationship discovery within MOCCs will help adapt to learner diversity where students from all over the globe take classes from MOOCs. Developing a fine-grained map of the concepts presented in the MOOC, indicating pre-requisite relationships, can facilitate students browsing into course materials flexibly. In addition, such a map can help in emphasizing the important topics in the course and how they are related, which can help improve students understanding. It can be further used to represent the knowledge state of a student at the concept level, and thus enable personalization in recommending course materials or quiz questions to students. In this paper, our goal is to construct such a map automatically for any course in order to accommodate students' diversity by supporting personalized learning.

Generating such a concept dependency graph presents a number of challenges. First, the instructor cannot predict the prior preparation of the students taking the class or the granularity at which she should develop the concept graph, and ensuring that such a concept graph remains up to date every year is time consuming. Second, an instructor does not introduce concepts in a rigid order, wherein she will always present the prerequisite concept before introducing the main concept; which makes it difficult in determining the presence and the direction of a relationship between concepts.

We propose innovative unsupervised methods to discover a directed concept dependency graph. We use lecture transcripts, as do Chaplot and Koedinger [2], to model the dependency structure between course concepts. Where Chaplot and Koedinger focus on modeling the prerequisite structure between units or lectures, we instead focus on inferring the dependency structure among concepts that appear *within and between* lectures. Our main technical innovation lies in exploiting the temporal ordering amongst concepts to discover the graph. To the best of our knowledge, we are the first to use temporal features to construct the dependency graph. We propose two measures—the Bridge Ensemble Measure and the Global Direction Measure—to infer the existence and the direction of the dependency relations between concepts.Both proposed measures outperform the baseline method [2] in AUC and the precision/recall curves.

The rest of the paper is organized as follows. In section 2, we formally frame our problem before describing the two proposed measures in section 3. Section 4 elaborates our approach for the evaluation and section 5 presents some limitations. Finally, we discuss some related work in section 6 before concluding our work on section 7.

2. PROBLEM DEFINITION

Informally, the problem explored in this work can be stated as follows: given course data, predict the dependency relationships between the course concepts. More formally, let X be the course represented by an ordered list of transcripts corresponding to each lecture: $X = [T_1, T_2, \dots, T_M]$ where M is the total number of lectures. Let C_X be the set of concepts discussed in the course $C_X = \{c_1, c_2, \dots, c_N\}$, where N is the total number of unique concepts. Given X and C_X, we aim to generate the concept dependency graph that relates concepts in C_X according to their prerequisite relationships. The resulting concept dependency graph is described by an edge weight matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$. Each entry a_{ij} of matrix $\mathbf{A}$ will contain the edge weight for the associated relationship $c_i \to c_j$, which means concept c_i is a prerequisite for concept c_j. The edge weight reflects the level of confidence in the inferred relationship. Notice that since the prerequisite relationship has a direction, $\mathbf{A}$ is **not** symmetric.

$$\mathbf{A} = \begin{bmatrix} 0 & \dots & \dots & W(c_1 \to c_N) \\ W(c_2 \to c_1) & 0 & \dots & W(c_2 \to c_N) \\ \dots & \dots & \dots & \dots \\ W(c_N \to c_1) & W(c_N \to c_2) & \dots & 0 \end{bmatrix}$$

The problem of constructing the concept dependency graph can be reduced to the problem of computing the edge weight between pairs of concepts given course data.

3. LINKING COURSE CONCEPTS

To relate the course concepts according to their dependency relationships, we propose two measures: the Bridge Ensemble Measure and the Global Direction Measure.

3.1 Bridge Ensemble Measure

The Bridge Ensemble Measure (BEM) captures concept dependency structure utilizing inter-lecture and intra-lecture strategies. It contains three components: Bridges, Sliding Windows, and the First Lecture Indicator.

3.1.1 Bridges

Let us look at how instructors naturally introduce concepts and their prerequisite(s). Let C_X be the set of concepts presented in course X and let c_a and c_b be concepts in that set. Determining the presence of a concept c_a in a lecture transcript T_i is discussed further in section 4.1. Suppose that c_a is a prerequisite to c_b. Then it stands to reason that (1) c_a will be introduced before c_b in the course progression, and (2) while explaining or talking about c_b, the instructor will naturally refer to c_a.

Bridge concepts allow us to exploit the temporal nature of lectures to infer concept dependency relationships across lectures. Intuitively, bridge concepts are introduced in an earlier lecture but re-appear in a later lecture when some new concept(s) are introduced. Accordingly, bridge concepts signal a prerequisite relationship from the bridge concepts to the new concepts introduced in the later lecture. For example, in Figure 1, the bridge concepts c_3 and c_4 are more likely to be prerequisite to concepts c_5, c_6, and c_7 discussed in lecture L_2. Formally, let L_i be the set of concepts in the lecture i in course X,

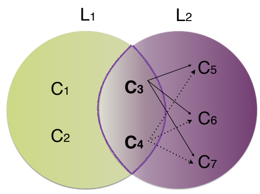

Figure 1: *The bridging concepts (c_3 and c_4) between lecture L_1 and L_2 and the resulting candidate prerequisite relationships.*

and L_j be the set of concepts for the lecture j where $j > i$. The intersection $L_j \cap L_i$ contains all the concepts that appear in both lectures. We call these **bridge concepts**. The difference $L_j \setminus L_i$ contains **difference concepts** which are the concepts present in the later lecture j but not in the earlier lecture i. If c_a belongs to the bridge concepts and c_b belongs to the difference concepts, then there is evidence for the dependency relationship $c_a \to c_b$ and the edge weight $W(c_a \to c_b)$ should increase. As a result, the bridge set $B_{ji} = \{(c_a \to c_b) \mid c_a \in L_j \cap L_i \wedge c_b \in L_j \setminus L_i\}$ contains all candidate prerequisite edges from lecture L_i to lecture L_j. If we replicate this exercise for every possible pair of lectures, we will end up with a comprehensive set of all possible candidate bridge edges **Bridges** for the course:

$$\mathbf{Bridges} = B_{M(M-1)} \cup B_{M(M-2)} \cup \dots \cup B_{21} \quad (1)$$

To calculate the edge weight of candidate edges in **Bridges**, we use the following bridge scoring function

$$W(c_a \to c_b) \approx F_{\mathbf{Bridges}}(c_a \to c_b) \quad (2)$$

where

$$\begin{aligned} & F_{\mathbf{Bridges}}(c_a \to c_b) \\ = & \frac{\text{The number of lectures where we observe both } c_a \text{ and } c_b}{\text{The number of lectures where we observe } c_b} \\ & = \frac{|\{L_j | c_a, c_b \in L_j\}|}{|\{L_j | c_b \in L_j\}|}. \end{aligned} \quad (3)$$

Keep in mind that the bridge scoring function will only calculated for candidate edges belong to **Bridges**. Other pairs of concepts will have zero value for the bridging score.

3.1.2 Sliding Windows

Bridge edges determined by the Bridge Method do not capture every possible prerequisite relationship. Consider the case where concept c_b has a strong prerequisite c_a, but c_a and c_b only appear together either in the set of bridge concepts ($L_j \cap L_i$) or in the set of difference concepts ($L_j \setminus L_i$). As a result, $c_a \to c_b$ will never appear in **Bridges** and hence the Bridge method cannot infer the prerequisite relations between them.

To solve this problem and capture intra-lecture prerequisite relationships, we zoom into each lecture and consider the proximity of concepts being presented in the lecture. Let $\vec{L_j} = [c_1, c_2, ..., c_n]$ be an ordered list of concepts discussed in lecture j, where n is the total number of concepts. Keep in mind that this ordered list contains redundant concepts which appear in the order where the instructor mentioned them. In the sliding windows method, we segment $\vec{L_j}$

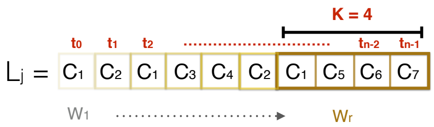

Figure 2: *A visualization example of lecture $\vec{L_j}$ with $r = n - K + 1$ sliding windows of size $K = 4$. The sliding windows captures the proximity of concepts.*

into windows $W_i = [c_i, \ldots, c_{i+K-1}]$ as follows:

$$\textbf{Windows}_j = \begin{cases} \{W_i \mid 1 \leq i \leq n-K+1\} & n \geq K \\ \{\vec{L_j}\} & n < K. \end{cases} \tag{4}$$

Figure 2 depicts the representation of lecture $\vec{L_j}$ using $r = n - K + 1$ windows of size $K = 4$. In this study, we choose the K that gives the best performance; K is set to *10* concepts.

The more windows in which c_a and c_b appear together, the stronger the relationship between c_a and c_b is; thus the edge weight should increase. The second component of the BEM for edge weights is the probability of the edge $c_a \rightarrow c_b$ given the information we have about all windows in all lectures $\textbf{Windows} = \bigcup_j \textbf{Windows}_j$.

$$W(c_a \rightarrow c_b) \approx F_{\textbf{Bridges}}(c_a \rightarrow c_b) + F_{\textbf{Windows}}(c_a \rightarrow c_b) \tag{5}$$

Where:

$$\begin{aligned} &F_{\textbf{Windows}}(c_a \rightarrow c_b) \\ &= \frac{\text{The number of windows where we observe } c_a \text{ and } c_b \text{ together}}{\text{The number of windows where we observe } c_b} \\ &= \frac{|\{W_i \in \textbf{Windows} \mid c_a, c_b \in W_i\}|}{|\{W_i \in \textbf{Windows} \mid c_b \in W_i\}|} \end{aligned} \tag{6}$$

We choose to accumulate the bridge weight with the sliding windows weight because these methods complement each other. Some edges that captured by the sliding windows method have zero bridging score and vice versa. Multiplying these two components instead of accumulating them would eliminate their effect in capturing inter- and intra-lecture prerequisite edges as the value of these edges will be zero.

3.1.3 First Lecture Indicator

The third component of the BEM for edge weights comes from the intuition that the context (other observed concepts) in which a new concept c_b is first introduced plays a strong role in determining what the prerequisite concepts of c_b are. We will assume that c_b is first introduced in lecture j when it has the highest term frequency of the concept c_b compared to other lectures. We call j the lecture indicator of c_b and denote it by $LI(c_b)$. When concept c_a appears in the lecture indicator of c_b ($c_a \in L_{LI(c_b)}$), then c_a might be a prerequisite to c_b. Another condition we need to examine is the temporal order of the lecture indicator of concept c_a. Naturally, when the instructor discusses a new concept, he or she needs to explain its prerequisite concepts beforehand, either in earlier lectures or in the same lecture where the new concept is being introduced. More formally, then, $LI(c_a) \leq LI(c_b)$. Thus when calculating $W(c_a \rightarrow c_b)$ we consider the first lecture indicator variable FLI_{c_a,c_b} where:

$$FLI_{c_a,c_b} = \begin{cases} 1, & \text{if } c_a \in L_{LI(c_b)} \text{ and } LI(c_a) \leq LI(c_b) \\ 0, & \text{otherwise} \end{cases}$$

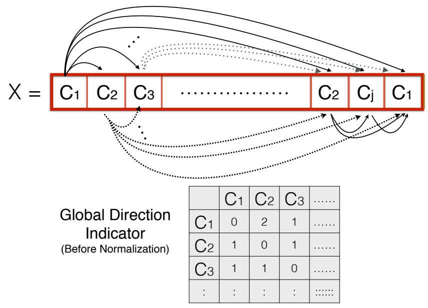

Figure 3: *A visualizing explanation of the Global Direction Indicator. X represents the course. The matrix contains the Global Direction Indicator (Before the normalization). Each element in the matrix represents how many times the concept c_{row} appears before the concept c_{column} in the whole entire course.*

The BEM for edge weights now becomes:

$$\begin{aligned} W(c_a \rightarrow c_b) &\approx F_{\textbf{Bridges}}(c_a \rightarrow c_b) \\ &+ F_{\textbf{Windows}}(c_a \rightarrow c_b) + FLI_{c_a,c_b} \end{aligned} \tag{7}$$

3.2 Global Direction Measure

The Global Direction Measure (GDM) is an alternative measure we propose to capture the dependency relationships between course concepts by incorporating time directly to consider not only the time ordering within lectures but also globally throughout the course delivery. In the Bridge Ensemble Measure, one problem with the sliding windows method is that the temporal order of concepts *within* a window W_i is ignored. This seems reasonable since in a single window, the instructor might mention the dependent concept before the prerequisite concepts. However, utilizing the temporal order of concepts in the entire course might improve the inference of the direction of the dependency relation. Thus, we propose the idea of the Global Direction Indicator (GDI).

The global direction indicator keeps track of the global temporal order frequency of concepts discussed in the course. In other words, it captures how many times concept c_a appears before concept c_b in the whole entire course. The more the concept c_a appears before the concept c_b, the more likelihood that the direction of the prerequisite relation is from c_a to c_b ($c_a \rightarrow c_b$). To capture the global direction indicator, we represent the course X as an ordered list of concepts discussed in all course lectures: $\vec{X} = [c_{11}, c_{12}, ..., c_{ij}, ..., c_{M1}, c_{M2}, ...]$ where i is the lecture number, j is the concept number, and M is the total number of lectures. Then, we keep track of temporal order frequency between any pair of concepts in the whole entire course. Figure 3 depicts the idea of the global direction indicator.

The formula of the global direction indicator is as follow:

$$GDI(c_a, c_b) = \frac{TOF(c_a \rightarrow c_b)}{\sum_{c_i \in C_X} TOF(c_a \rightarrow c_i)} \tag{8}$$

where TOF is the temporal order frequency, c_i are all concepts appear after c_a in the course progression. We normalize the TOF of $c_a \rightarrow c_b$ by the total number of times c_a appears before any other concept in the course to reduce the impact of popular concepts that tend to appear before almost every other concept in the course.

In addition to the global direction indicator, we modify the sliding

windows method to consider the local temporal order of concepts within a single window:

$$
\begin{aligned}
&F_{\textbf{Dir-Windows}}(c_a \to c_b) \\
&= \frac{\text{The number of windows where we observe } c_a \to c_b}{\text{The number of windows where we observe } c_b} \\
&= \frac{|\{W_i \in \textbf{Windows} \mid c_a \to c_b \in W_i\}|}{|\{W_i \in \textbf{Windows} \mid c_b \in W_i\}|} \qquad (9)
\end{aligned}
$$

In this case, the directed sliding windows (Dir-Windows) method captures not only the proximity of pair of concepts but also the local direction within lectures while the global direction indicator captures the frequency of the global direction.

The edge weight function according to the GDM is as follow:

$$W(c_a \to c_b) \approx GDI(c_a, c_b) \times F_{\textbf{Dir-Windows}}(c_a \to c_b) \qquad (10)$$

The rationale behind combining the GDM Components by multiplying them instead of accumulating them is to use the global direction indicator to improve the direction of edges predicted by the directed windows instead of predicting the existence of edges. The problem with the global direction indicator in predicting the edge existence is that it might give high weight to concepts that appear very often with the same direction order even if they do not appear together in any lecture.

4. EVALUATION

In this section, we demonstrate the evaluation process conducted to assess the performance of the proposed measures. We utilize the course "Text Retrieval and Search Engines"[1] to construct the concept dependency graph to evaluate our developed measures.

4.1 Building the Course Concept Space

The focus of our work is on understanding how to infer the dependency relationship between concepts, but in order to evaluate the proposed measures, we must first construct a set of concepts. There is a wide body of work which attempts to solve the problem of defining and inferring concepts [3, 9, 10]. In this paper, we use a pre-trained part-of-speech-guided phrasal segmentation, called Autophrase [10, 8], to extract salient phrases from lectures' transcripts. While Autophrase generates many good salient phrases, some phrases are either too general or are verb phrases. Our approach to improve the quality of the selected phrases is to extract phrases from weekly overviews using the same phrasal segmentation method. At the beginning of each week in the course, there is a week overview page that explains the goals and objectives of that week along with the key phrases and concepts that students need to understand. Utilizing the overview page of each week aids in filtering out meaningless phrases.

After extracting salient phrases, we manually group synonym phrases together to construct a concept. We follow Siddiqui et al. [11] definition of concepts by defining a concept as a set of salient phrases that describe it. This design decision was made to allow for flexibility in concept description since the same concept can be referred to using different phrases by different people.

4.2 Ground Truth

To evaluate the effectiveness of the proposed measures, we form a ground truth concept graph by leveraging students submissions

[1] `https://www.coursera.org/learn/text-retrieval`

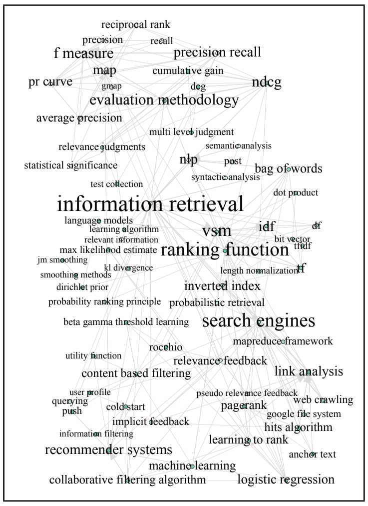

Figure 4: *The visualization of the ground truth graph generated from students data.*

about concept dependencies in a course (CS 410) at UIUC offered in the Spring 2017 semester that follows Coursera's "Text Retrieval and Search Engines." Students were asked to submit a weekly summary of new concepts they have learned along with prerequisite concepts. The following is an example of a student entry from week 3:

```
# f-measure: precision, recall
# pr curve: precision, recall
# map: arithmetic mean of average precision
# gmap: geometric mean of average precision
```

The total number of edges in the ground truth were 239 edges for 74 concepts in the concept space. Figure 4 visualizes the ground truth concept graph to see how concepts are related. It is clear that concepts such as "information retrieval", "search engines", "ranking function", and "evaluation methodology" have higher degree as these concepts are connected with many other concepts in the course. This is reasonable as these concepts considered fundamental in this course. Such a figure can also be seen as a useful topic map that can facilitate students browsing into course materials covering different topics flexibly; however, the map shown in this figure was constructed based on student submissions—with the proposed methods, we can construct such a map automatically for any course.

4.3 Baseline Approach

Since the problem formulation of using only transcripts to predict concept dependency is novel, strictly speaking, no previous method can be directly used to produce the desired output. The closest work that we can compare with is the work of Chaplot and Koedinger [2], which also only uses course content without any external knowledge. In their paper, they develop two methods: a text-based method called the *overlap method*, and a performance-based method. Since our work is a text-based method, we compare our measures to the overlap method. The main difference between our work and the overlap method is that we exploit the temporal features of course delivery while the overlap method does not; this makes the overlap method an ideal baseline to study the effect of the temporal features on the accuracy of edge prediction.

The overlap method, however, only predicts the prerequisite relations

Table 1: *Performance (area under ROC curve) of concept graph generation for the three methods considered. Both of the new measures introduced in the paper outperform the state-of-the-art ExtendedOverlap method on both edge existence and edge direction tasks.*

	AUC (ROC)	
Method	Existence	Direction
Bridge Ensemble Measure	**0.80**	**0.81**
Global Direction Measure	**0.80**	**0.78**
ExtendedOverlap Method	0.74	0.74

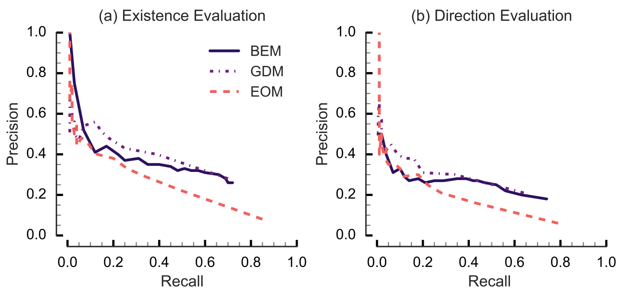

Figure 5: *The Precision/Recall Curves of Bridge Ensemble Measure (BEM), Global Direction Measure (GDM), and the baseline ExtendedOverlap Method (EOM). GDM and BEM outperform the baseline method (EOM) in both the existence evaluation and direction evaluation.*

between units (e.g. lectures) using the text overlap between units. Thus the method cannot be used directly to predict dependency between concepts, the problem that we attempt to solve. Therefore, we propose an extension called **ExtendedOverlap** for solving our problem as a baseline for comparison. Our main idea for extending the overlap method is to first map a course to a set of lectures where the concept occurred and then leverage the lecture dependency relations predicted using the overlap method to assess the dependency between two concepts by accumulating the weight of the dependency relations of lectures they belong to. All weights are normalized to be between zero and one. We implemented the overlap method using the noun phrases with document frequency normalization since they achieve the highest performance [2].

4.4 Concept Graph Performance

We conduct the evaluation of the performance of the generated concept graphs over two dimensions: edge existence and edge direction. Edge existence evaluates whether the method predicts correct edges or not while edge direction evaluation ensures not only the correctness of the edge prediction but also their direction. The AUC values of all the methods are shown in Table 1. We can notice that both the Bridge Ensemble Methods (BEM) and Global Direction Measure (GDM) outperform the baseline ExtendedOverlap (EOM) in terms of the AUC values for both the existence task and the direction task.

We also use the precision/recall curve to compare various methods as shown in Figure 5. It appears that the Global Direction Measure has the highest curve followed by the Bridge Ensemble Measure in both dimensions. This indicates that for various recall values our measures predict more accurate edges than the baseline. It is

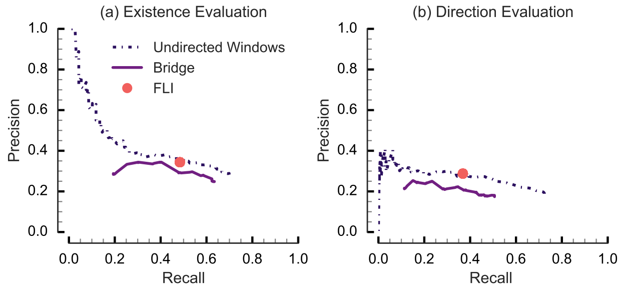

Figure 6: *The comparison between the performance of the Bridge Ensemble Measure components. While the undirected sliding windows correctly captured the edge existence in the interval [0.0, 0.2], it fails at predicting edge directions.*

also interesting to notice that in the precision/recall curve of the existence evaluation (Figure 5 (a)), the Bridge Ensemble Measure has the highest precision when the recall is less than **0.1** while in the precision/recall curve of the direction evaluation (Figure 5 (b)) has the lowest precision until it reaches the recall value of **0.2**. This indicates that, in the interval **[0.0,0.2]**, the Bridge Ensemble method captures the existence of the edges but fails at specifying the correct direction. To examine the reason, we study the performance of various components of the Bridge Ensemble Measure as depicted in Figure 6. It is appear that the undirected sliding windows method has the highest curve in the existence evaluation (Figure 6 (a)) and since it only captures the proximity of pair of concepts and how they are related, it surges the precision/recall curve of the existence performance in the interval **[0.0,0.2]** by capturing correct prerequisite edges. However, since the temporal feature is only used in limited way as a binary variable among lectures through bridges and first lecture indicator components, it sometimes fails at predicting the correct direction of edges between concepts that only appears within the same lectures. In contrast, the Global Direction Measure exploits the global direction indicator that keeps track of the global temporal order frequency and hence emphasizes or corrects the direction captured by the directed sliding windows method as depicted in Figure 7. It is clear from Figure 7 that the global direction indicator improves the edge direction of the directed windows method when the recall value is less than **0.2** while it emphasize the edge direction of the directed windows after that.

To further analyze the differences between the Bridge Ensemble Measure and the Global Direction Measure, we examine their behavior in the existence dimension. We found that all true positive edges and false positive edges captured by Global Direction measure are also captured by Bridge Ensemble Measure. However, Bridge Ensemble Measure has more false positive edges (59 edges) and more true positive edges (only 4 edges). We examine the source of the extra false positive edges in the Bridge Ensemble Measure and found that 73% came from the bridge method, 3% came from the first lecture indicator, and 22% are from both the bridge method and the first lecture indicator while the sliding windows has zero contribution (0%). Further examination of these extra false positive errors shows that some of them capture long distance dependencies such as the relation “natural language processing” → “recommender systems”, which captures the dependency between the concepts explained in the first and last lectures. By examining the source of this relation, we found that the bridge method makes the inference of the relation. As

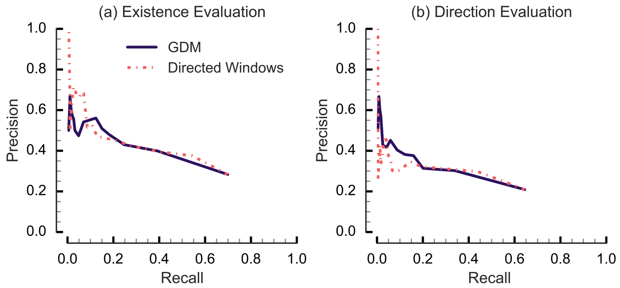

Figure 7: *The effect of the global direction indicator on the Global Direction Measure. The GDI improves the edge direction of the directed windows method when the recall value is less than **0.2** while it emphasize the edge direction of the directed windows after that*

mentioned earlier, bridge method captures the dependency relations between concepts across lectures and, in contrast to the sliding windows method, it does not require the proximity of concepts within lectures' transcripts. This property of the bridge method gives the Bridge Ensemble Measure the ability to capture long distance relations between concepts in contrast to the Global Direction Measure which only captures the local dependencies between concepts (within lectures).

We also conduct a qualitative analysis of the false positive edges to examine the reason of the high values and hence the low precision values. We found three types of false positive edges that we may actually consider correct relations. First, the transitive property edges that are captured by our measures are not always specified in the ground truth edges. For example, students specify the relations "length normalization" → "ranking function", and "ranking function" → "vector space model". While both our measures and the baseline capture these relations, they go further and also capture the transitive relation "length normalization" → "vector space model". Second, there are issues with relations with differing concept granularities. For instance, students specify a dependency relation "language models" → "dirichlet prior smoothing" while the generated graphs by the three methods capture the relation "language models" → "smoothing methods." The concept "smoothing methods" is more general than the concept "dirichlet prior smoothing." Third, there are missing "true" relations that the students did not specify in the ground truth. For example, students did not specify the following relations that are captured by our measures: "tfidf" → "bm25", and "length normalization" → "bm25." In general, the three types of false positive errors can justify to some extent the high values of the false positive errors and thus the low values of the precision.

In general, the Bridge Ensemble Measure and the Global Direction Measure outperform the baseline in terms of AUC and precision/recall curves, with the Global Direction Measure having the overall highest performance. These results emphasize the positive effect of the temporal feature on improving the accuracy of the generated concept graph.

5. LIMITATIONS

There are some limitations in our study. First, in the evaluation we have not examined the robustness of our measures compared to the baseline utilizing other courses taught by different instructors. Second, we use the students' perspectives of the concept dependency graph as a ground truth, and we are the first study to do so. However, in the future we plan to compare various methods' performance by utilizing not only the students' perspectives of the concept graph but also one generated by instructors. Third, in this study, we include an edge in the ground truth even if only one student specifies it; in the future we plan to use some agreement measures before including an edge in the ground truth. Fourth, we represent the course concept graph according to the dependency structure without distinguishing whether the dependency relation captures the hierarchical structure or real prerequisite relationships. We believe that the ideal structure of the concept dependency graph is a hierarchical graph with cross link edges where the hierarchical structure captures the "general concept" to "specific concept" relations while the cross links depict the prerequisite relationships between concepts.

6. RELATED WORK

Most prior work focuses on relationships between concepts such as similarity relations [13] and hierarchical relations [5]. Although the most important concept relation to learners is the dependency or prerequisite relation, this relation has been the least studied [4]. Some prior works utilize Wikipedia articles [6, 12, 1, 7], scientific corpora [4], or educational materials from online educational platforms [14, 2, 7] to model the dependency structure between concepts. While many studies utilized external knowledge to recover the prerequisite relations [14, 7] , Chaplot and Koedinger [2] utilize the course content with students performance to infer such relation. In contrast, to make our method more accessible, we exploit only the easily accessible educational materials to model the dependency relations among course concepts.

Previous research represents graph concepts in various ways. Gordon et al. [4] identify concepts using LDA topic modeling that fails in identifying finer-grained concepts. Yang et al. [14] explored four different representations and found that word and category representations have similar performance; however, word representation has slightly better performance on some data sets. One problem with using category representations is that mapping phrases to Wikipedia categories affects concept granularities by preferring more general concepts. On the other hand, Chaplot and Koedinger [2] found that noun phrase representation outperforms other representations. Therefore, in this study, we utilize noun phrase representation but extend it using temporal information.

Previous work developed supervised [1, 12, 14] and unsupervised approaches [6, 7, 2] to predict the dependency relationships among concepts. Several studies rely on external knowledge to predict prerequisite relations across courses [14, 7] while we only leverage course materials to model the dependency relations within a course not between courses. Chaplot and Koedinger [2] address the dependency structure within courses, but between units instead of concepts taught within units. Another main difference is the use of the temporal feature in the course delivery to model the dependency structure as we are the first study that exploits the temporal feature.

7. CONCLUSIONS

In this paper, we leverage the accessible MOOC content and incorporate the temporal feature of the course to construct a concept dependency graph. We developed Bridge Ensemble Measure and Global Direction Measure that exploit the temporal order in course delivery to model the dependency structure. We revealed in the evaluation that both developed measures outperform the baseline method in AUC and in precision recall curves. This finding emphasizes the positive effect of utilizing the temporal feature of course progression.

8. REFERENCES

[1] R. Agrawal, B. Golshan, and E. Papalexakis. Toward data-driven design of educational courses: A feasibility study. *JEDM-Journal of Educational Data Mining*, 8(1):1–21, 2016.

[2] D. Chaplot and K. R. Koedinger. Data-driven automated induction of prerequisite structure graphs. In *Proceedings of the 9th International Conference on Educational Data Mining*, pages 318–323. EDM, 2016.

[3] A. El-Kishky, Y. Song, C. Wang, C. R. Voss, and J. Han. Scalable topical phrase mining from text corpora. *Proceedings of the VLDB Endowment*, 8(3):305–316, 2014.

[4] J. Gordon, L. Zhu, A. Galstyan, P. Natarajan, and G. Burns. Modeling concept dependencies in a scientific corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 866–875, 2016.

[5] I. Jonyer, D. J. Cook, and L. B. Holder. Graph-based hierarchical conceptual clustering. *Journal of Machine Learning Research*, 2(Oct):19–43, 2001.

[6] C. Liang, Z. Wu, W. Huang, and C. L. Giles. Measuring prerequisite relations among concepts. In *EMNLP*, pages 1668–1674, 2015.

[7] C. Liang, J. Ye, Z. Wu, B. Pursel, and C. L. Giles. Recovering concept prerequisite relations from university course dependencies. In *AAAI*, pages 4786–4791, 2017.

[8] J. Liu, L. Jiang, Z. Wu, Q. Zheng, and Y. Qian. Mining learning-dependency between knowledge units from text. *The VLDB Journal-The International Journal on Very Large Data Bases*, 20(3):335–345, 2011.

[9] J. Liu, J. Shang, C. Wang, X. Ren, and J. Han. Mining quality phrases from massive text corpora. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1729–1744. ACM, 2015.

[10] J. Shang, J. Liu, M. Jiang, X. Ren, C. R. Voss, and J. Han. Automated phrase mining from massive text corpora. *arXiv preprint arXiv:1702.04457*, 2017.

[11] T. Siddiqui, X. Ren, A. Parameswaran, and J. Han. Facetgist: Collective extraction of document facets in large technical corpora. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 871–880. ACM, 2016.

[12] P. P. Talukdar and W. W. Cohen. Crowdsourced comprehension: predicting prerequisite structure in wikipedia. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pages 307–315. Association for Computational Linguistics, 2012.

[13] R. W. White and J. M. Jose. A study of topic similarity measures. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 520–521. ACM, 2004.

[14] Y. Yang, H. Liu, J. Carbonell, and W. Ma. Concept graph learning from educational data. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pages 159–168. ACM, 2015.